A Two-Stage Bayesian Variable Selection Method with the Extension of Lasso for Geo-Referenced Data

Authors : Georgiana Onicescu, Yuqian Shen

Abstract : Due to the complex nature of geo-referenced data, multicollinearity of the risk factors in public health spatial studies is a commonly encountered issue, which leads to low parameter estimation accuracy because it inflates the variance in the regression analysis. To address this issue, we proposed a two-stage variable selection method by extending the least absolute shrinkage and selection operator (Lasso) to the Bayesian spatial setting, investigating the impact of risk factors to health outcomes. Specifically, in stage I, we performed the variable selection using Bayesian Lasso and several other variable selection approaches. Then, in stage II, we performed the model selection with only the selected variables from stage I and compared again the methods. To evaluate the performance of the two-stage variable selection methods, we conducted a simulation study with different distributions for the risk factors, using geo-referenced count data as the outcome and Michigan as the research region. We considered the cases when all candidate risk factors are independently normally distributed, or follow a multivariate normal distribution with different correlation levels. Two other Bayesian variable selection methods, Binary indicator, and the combination of Binary indicator and Lasso were considered and compared as alternative methods. The simulation results indicated that the proposed two-stage Bayesian Lasso variable selection method has the best performance for both independent and dependent cases considered. When compared with the one-stage approach, and the other two alternative methods, the two-stage Bayesian Lasso approach provides the highest estimation accuracy in all scenarios considered.

Keywords : Lasso, Bayesian analysis, spatial analysis, variable selection
Conference Title : ICBB 2020 : International Conference on Biometrics and Biostatistics
Conference Location : Paris, France
Conference Dates : June 25-26, 2020